Journal of Electronics, Electromedical Engineering, and Medical Informatics
Homepage: jeeemi.org; Vol. 8, No. 1, January 2026, pp: 270-283 e-ISSN: 2656-8632

RESEARCH ARTICLE OPEN ACCESS

Optimizing Input Window Length and Feature Requirements for Machine Learning-Based Postprandial Hyperglycemia Prediction

Muhammad Rafly Alfarizqy Maulana, Fatma Indriani, Friska Abadi, Dwi Kartini, and Muhammad Itqan Mazdadi

Department of Computer Science, Lambung Mangkurat University, Banjarbaru, Indonesia

Corresponding author: Fatma Indriani (e-mail: f.indriani@ulm.ac.id), **Author(s) Email**: Muhammad Rafly Alfarizqy Maulana (e-mail: rafalfarizqy.ra@gmail.com), Friska Abadi (e-mail: friska.abadi@ulm.ac.id), Dwi Kartini (e-mail: dwikartini@ulm.ac.id), Muhammad Itqan Mazdadi (e-mail: mazdadi@ulm.ac.id)

Abstract **Continuous glucose monitoring systems currently generate alerts only after blood glucose thresholds are breached, limiting their utility for proactive diabetes management. Predicting postprandial glucose excursions before they occur requires determining the optimal amount of historical data and identifying which features contribute most to prediction accuracy. This study systematically evaluates how the length of the pre-meal observation window and feature composition affect machine-learning predictions of hyperglycemia events 60 minutes after eating. We analyzed 1,642 meal events from 45 adults wearing continuous glucose sensors, constructing features from pre-meal glucose trajectories, meal macronutrients, time of day, and health status. Four observation windows (15, 30, 45, 60 minutes) and three feature sets (all features, glucose-only, meal-only) were evaluated using Random Forest, XGBoost, and CatBoost with 5-fold group cross-validation. CatBoost with a 30-minute window achieved the best performance: 72.6% F1-macro, 79.6% accuracy, and 64.0% recall for hyperglycemia detection. Extending windows beyond 30 minutes did not yield consistent benefits, whereas 15-minute windows yielded comparable results. Glucose trajectory features alone retained 94% of full model performance (68.5% F1-macro), whereas meal composition alone proved insufficient (59.4% F1-macro). These findings demonstrate that recent glucose history dominates short-term prediction, enabling practical real-time systems with minimal data requirements. A 30-minute observation window with glucose and meal features offers an effective balance between prediction accuracy and system responsiveness.**

Keywords **Hyperglycemia, CGM, Window Length, Feature Composition, CatBoost**.

I. Introduction

Diabetes affects over 537 million adults worldwide, with projections reaching 783 million by 2045 [1]. Elevated postprandial blood glucose contributes significantly to long-term complications [2][3], yet current monitoring approaches remain primarily reactive. Continuous glucose monitoring (CGM) technology samples interstitial glucose every 5 minutes, providing detailed trajectories of daily fluctuations [4][5]. However, commercial CGM systems typically alert users only after glucose exceeds predefined thresholds or when rates of change become concerning [6]. The inherent 5-15-minute lag between blood and interstitial glucose measurements further delays these alerts, often arriving after problematic glucose excursions have already begun [7].

Machine learning offers a path toward predictive rather than reactive glucose management [8][9]. Gradient boosted decision tree algorithms, including XGBoost, CatBoost, and LightGBM, have demonstrated strong performance on structured biomedical data. Specifically, these tree-based ensembles are well-suited for this domain due to their ability to handle nonlinear relationships between glucose features and meal timing without extensive data scaling, and their robustness to overfitting relative to deep neural networks on tabular datasets with limited sample sizes [10][11][12]. The success of these methods in diabetes prediction has been attributed to their ability to handle missing data gracefully and to provide interpretable feature-importance rankings [13][14]. Several studies have shown that combining CGM data with meal composition information can improve the accuracy of predictions of postprandial glucose responses [15][16]. These capabilities suggest the potential to build systems that warn users of impending glucose excursions before planned meals, enabling informed food choices or preemptive interventions. Despite recent progress, two fundamental design questions remain inadequately

Manuscript 10 November 2025; Revised 5 January 2026; Accepted 12 January 2026; Available online 17 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1401

addressed. First, the optimal length of historical CGM data needed for reliable prediction remains an open research question. While prior studies have utilized observation windows ranging from 30 minutes to several hours [17][18], these selections are often arbitrary or strictly dataset-dependent. There is a lack of rigorous comparative frameworks that systematically evaluate the trade-off between observation duration and predictive accuracy within the same experimental setting. Consequently, it remains unclear whether longer windows genuinely improve model generalization or merely increase computational complexity and latency. While shorter windows reduce system latency and storage needs, they risk missing important pre-meal trends. Therefore, establishing the minimal effective history that preserves predictive performance, particularly for the clinically relevant task of predicting events 60 minutes after eating, is crucial for developing responsive real-time systems, a gap that this study specifically aims to address.

Second, the relative contribution of different data modalities, specifically the marginal benefit of adding detailed macronutrient data to recent glucose trends, remains debated in the literature, often due to a lack of comparative frameworks. While nutritional science establishes clear relationships between macronutrients and glucose response, computational studies report mixed results regarding whether meal composition features improve prediction beyond glucose trajectories alone [8][13][11]. The mixed results regarding meal composition in the current literature often stem from a lack of rigorous ablation experiments. Many studies report performance gains by adding features but fail to quantify the marginal utility of each modality. Understanding which features drive predictive performance has direct implications for system design, data collection requirements, and deployment feasibility in resource-constrained settings.

Hyperglycemia prediction is inherently imbalanced, with normal postprandial outcomes substantially outnumbering hyperglycemic events. Without appropriate handling, machine learning models tend to favor the majority class, leading to poor sensitivity for detecting the clinically critical hyperglycemia events [19][20]. Cost-sensitive training and suitable validation schemes are therefore essential [21][22]. This research systematically addresses both design questions through controlled ablation experiments. We evaluate four pre-meal observation windows (15, 30, 45, and 60 minutes) to determine the minimal effective history length for 60-minute-ahead prediction. We compare three feature configurations (all features, glucose-only, meal-only) to quantify the contribution of glucose trajectories versus meal composition. Three gradient-boosting algorithms (Random Forest, XGBoost, CatBoost) are trained on 1,642 meal events from 45 participants, using rigorous group cross-validation to assess generalization to unseen individuals. Class weighting addresses the inherent imbalance between normal and hyperglycemic outcomes without synthetic data generation.

This work makes three contributions to applied machine learning for glucose prediction. First, to provide empirical evidence on the minimal observation window required for reliable prediction, demonstrating that short pre-meal histories achieve performance comparable to extended observation periods. This enables low-latency prediction systems with reduced buffering requirements for real-time deployment. Second, through systematic feature ablation, we quantify the relative contributions of glucose trajectories and meal composition, revealing which data modalities drive predictive performance and which can be omitted without substantial loss of accuracy. Third, we evaluate design choices across three tree-based models using rigorous group cross-validation, ensuring that the findings reflect true generalization to unseen individuals and yielding practical guidance for building reliable, low-latency hyperglycemia prediction systems.

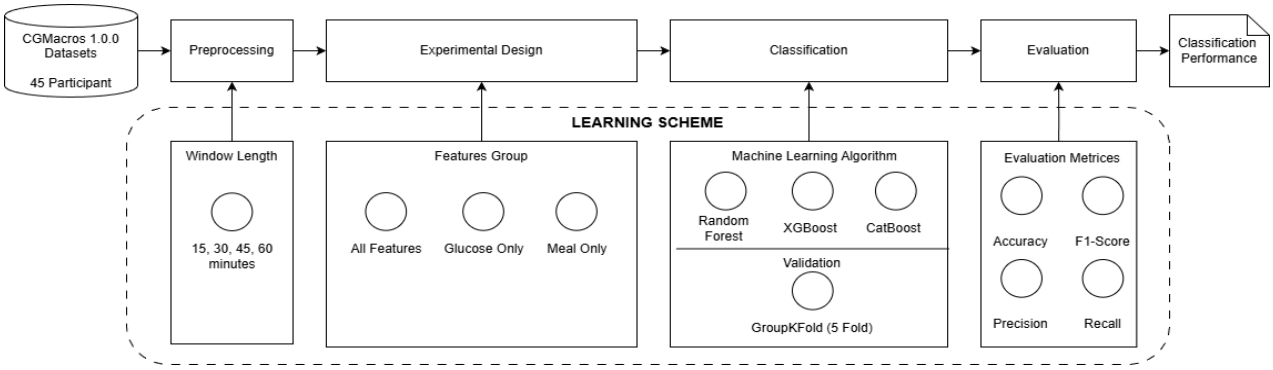

Fig. 1. The Research workflow of input window length and feature composition in postprandial hyperglycemia classification

Manuscript 10 November 2025; Revised 5 January 2026; Accepted 12 January 2026; Available online 17 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1401

II. Method

This study examined the impact of input window length and feature composition on the classification of postprandial hyperglycemia using machine learning approaches. The research process starts from dataset loading, data preprocessing, experimental design, classification using three algorithms, and evaluation using several metrics. The research flow is shown in Fig. 1.

A. Data Collection

This study utilized the CGMacros dataset (version 1.0.0), a publicly available multimodal dataset accessible via PhysioNet [23][24]. The complete dataset, along with its associated documentation, is publicly accessible at the following link: https://physionet.org/content/cgmacros/1.0.0/. This dataset was collected at the Sansum Diabetes Research Institute in Santa Barbara, California, and comprises data from 45 adults who were monitored continuously for 10 consecutive days in a free-living environment. Participants included 15 healthy individuals, 16 with prediabetes, and 14 with type 2 diabetes, as determined by HbA1c measurements during initial screening. Although this research utilizes a secondary public dataset, the original data collection was conducted under approved ethical protocols (IRB approval), and all participant records were fully de-identified to ensure privacy compliance before public release, adhering to the Health Insurance Portability and Accountability Act (HIPAA) Safe Harbor standards, negating the need for specific ethical approval for this secondary analysis.

Each participant wore two blinded continuous glucose monitoring devices: the Abbott FreeStyle Libre Pro (15-minute sampling interval) and the Dexcom G6 Pro (5-minute sampling interval). Participants also wore a Fitbit Sense activity tracker and logged all meals using a smartphone application. Meal documentation included photographs taken before and after consumption along with macronutrient estimates (carbohydrates, protein, fat, fiber, and calories). Breakfasts consisted of protein shakes with varying macronutrient compositions, lunches were ordered from a standardized restaurant chain (*Chipotle Mexican Grill*), and dinners comprised foods of the participant's own choice. To minimize interference from prior meals, participants were instructed to maintain at least 3 hours between consecutive meals and to consume only water or unsweetened coffee during those intervals.

For this study, we focused exclusively on Dexcom G6 data due to its higher temporal resolution (5-minute sampling). The dataset provides preprocessed tabular files in which individual meal episodes have been identified and features engineered with respoect to meal onset times. After quality control procedures that excluded episodes with missing CGM readings or incomplete meal annotations, the final dataset comprised 1,642 meal events from all 45 participants. Each meal episode was classified into one of two outcome categories based on the maximum glucose concentration measured within 60 minutes post-meal (h60): **normal** (glucose ≤180 mg/dL) or **hyperglycemia** (glucose >180 mg/dL). The observed class distribution was 75.4% normal (1,238 events) and 24.6% hyperglycemia (404 events), corresponding to a 3.06:1 imbalance ratio.

B. Feature Engineering and Input Windows

This study focused exclusively on Dexcom G6 CGM readings because they have a higher temporal resolution (5-minute sampling interval) compared to the Abbott FreeStyle Libre Pro (15-minute sampling). Each sample in the preprocessed dataset represents a single meal event with features engineered relative to meal onset time (t=0). The primary prediction objective is binary classification of postprandial hyperglycemia, defined as a maximum glucose concentration exceeding 180 mg/dL within 60 minutes post-meal (h60). This threshold aligns with clinical standards for postprandial hyperglycemia as defined by the American Diabetes Association [25], and resulted in an observed class distribution of 75.4% normal outcomes (1,238 events) and 24.6% hyperglycemia (404 events), corresponding to a 3.06:1 imbalance ratio.

To investigate the minimal temporal context required for reliable prediction, we evaluated four pre-meal observation windows: 15, 30, 45, and 60 minutes (denoted W15, W30, W45, W60). These specific intervals were chosen to capture distinct physiological phases of glucose dynamics: the 15-minute window captures the immediate rate of change, while 30-to 45-minute windows encompass the typical onset of insulin action. Extending to 60 minutes aligns with the standard postprandial horizon, allowing the model to observe the full pre-meal trend stability. Each window W of length w minutes contains glucose measurements sampled at 5-minute intervals leading up to meal onset (t=0) expressed in Eq. (1) [23]:

$$\mathrm{G}^{(\mathrm{w})} = \mathrm{GL}_{(\mathrm{t}-\mathrm{w})}, \mathrm{GL}_{(\mathrm{t}-\mathrm{w}+5)}, \ldots, \mathrm{GL}_{(\mathrm{t}-5)}, \mathrm{GL}_{\mathrm{t}}0 \quad (1)$$

where the number of lag features is $n = w/5 + 1$. Thus, W60 yields 13 historical measurements ($\mathrm{GL}_{t-60}, \mathrm{GL}_{t-55}, \ldots, \mathrm{GL}_{t-5}$) plus the current glucose reading at meal time (GL_{t0}), while W45, and W30, and W15 contain 10, 7, and 4 lag features, respectively. Correlation analysis revealed that recent glucose measurements exhibit the strongest association with hyperglycemia outcomes, with GL_{t0} showing the highest correlation (0.458), followed by GL_{t-5}(0.451) and GL_{t-10} (0.447). This monotonic decrease in correlation with temporal distance motivated our focus on compact observation windows.

Manuscript 10 November 2025; Revised 5 January 2026; Accepted 12 January 2026; Available online 17 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1401

Beyond raw glucose lags, each window incorporates two derived statistical features. The exponential moving average (*GL_EMA*) provides a smoothed representation of the glucose trajectory over the observation period, emphasizing recent measurements through exponential weighting. *GL_EMA* is computed iteratively as Eq. (2) [23]:

$$GL\_EMA_t = \alpha \cdot GL_t + (1-\alpha) \cdot GL\_EMA_{t-1} \quad (2)$$

where α = 0.3 is the smoothing parameter that controls the weight assigned to recent observations relative to historical values. The linear slope (*GL_slope*) captures the directional trend of glucose change through least-squares regression, as shown in Eq. (3) [23]:

$$GL\_slope = \frac{\sum(t_i-\bar{t})(GL_i-\overline{GL})}{\sum(t_i-\bar{t})^2} \quad (3)$$

where t_i represents time points within the observation window, and the barred variables denote mean values. Negative slope values indicate declining glucose, while positive values indicate rising glucose prior to the meal. Across the dataset, *GL_EMA* exhibited a mean of 126.6 mg/dL (SD=32.1), while *GL_slope* showed a slight average decline (-0.11 mg/dL per minute, SD=0.43), reflecting typical pre-meal glucose stabilization. Temporal context features encode meal timing through cyclical transformations and binary indicators. Hour of day is represented using sine and cosine functions to capture the circular nature of time as in Eq. (4) and Eq. (5) [23].

$$\text{hour\_sin} = \sin\left(\frac{2\pi \cdot \text{hour}}{24}\right) \quad (4)$$

$$\text{hour\_cos} = \cos\left(\frac{2\pi \cdot \text{hour}}{24}\right) \quad (5)$$

This encoding preserves the proximity between late evening (23:00) and early morning (01:00) hours, which linear hour encoding would misrepresent. Additionally, two binary meal-period indicators distinguish breakfast (is_morning: 06:00-11:59) and dinner (is_evening: 18:00-23:59) periods, capturing circadian variations in insulin sensitivity that are known to affect the postprandial glucose response.

Meal composition features quantify macronutrient content consumed at each meal: carbohydrates, protein, fat, fiber (all in grams), and total calories. These features remain constant across all window lengths as they describe the meal itself rather than pre-meal history. Observed macronutrient distributions showed substantial variability (carbohydrates: mean=51.7g, SD=40.6g; protein: mean=29.0g, SD=25.6g; fat: mean=19.2g, SD=20.7g), reflecting the diverse meal types consumed during the 10-day monitoring period. Regarding health status, participants' diagnostic status (Healthy, Prediabetes, Type 2 Diabetes) was encoded as a categorical variable. This feature accounts for baseline metabolic differences (e.g., insulin resistance levels) that fundamentally alter glucose trajectories distinct from meal-specific responses.

Table 1 summarizes the feature composition for each observation window configuration. Features were extracted from the preprocessed dataset with minimal transformation. Glucose-derived statistical features (EMA, slope) and cyclically-encoded temporal features were computed as specified in Eq. (2-5), while glucose lags and meal macronutrients retained their original scales. No normalization or standardization was applied. Columns that could leak information about the prediction target were systematically excluded, including participant identifiers, meal timestamps, and any glucose measurements beyond h60. This feature engineering approach balances simplicity with physiological relevance. The lag features capture glucose trajectory dynamics; statistical features summarize trends; temporal features account for circadian effects; meal features provide nutritional context; and demographic information reflects baseline metabolic state. The independent-window design enables direct comparison of observation-length

Table 1. Feature composition by observation window

Window	Glucose Lags	Statistical	Temporal	Meal	Demographic	Total Features
W15	GL_t-15 to GL_t0 (4)	GL_EMA, GL_slope (2)	hour_sin, hour_cos, is_morning, is_evening (4)	Carbs, Protein, Fat, Fiber, Calories (5)	Diagnosis (1)	15
W30	GL_t-30 to GL_t0 (7)	GL_EMA, GL_slope (2)	hour_sin, hour_cos, is_morning, is_evening (4)	Carbs, Protein, Fat, Fiber, Calories (5)	Diagnosis (1)	18
W45	GL_t-45 to GL_t0 (10)	GL_EMA, GL_slope (2)	hour_sin, hour_cos, is_morning, is_evening (4)	Carbs, Protein, Fat, Fiber, Calories (5)	Diagnosis (1)	21
W60	GL_t-60 to GL_t0 (13)	GL_EMA, GL_slope (2)	hour_sin, hour_cos, is_morning, is_evening (4)	Carbs, Protein, Fat, Fiber, Calories (5)	Diagnosis (1)	24

Manuscript 10 November 2025; Revised 5 January 2026; Accepted 12 January 2026; Available online 17 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1401

Table 2. Experimental design description

Experiment	Feature Groups Included	Windows Evaluated	Purpose
Experiment 1: All Features	Glucose + Meal + Temporal + Diagnosis	W15, W30, W45, W60	Baseline
Experiment 2: Glucose-Only	Glucose lags, EMA, slope only	W15, W30, W45, W60	Isolate the physiological signal
Experiment 3: Meal-Only	Meal macros + Temporal + Diagnosis	W15 (fixed)	No CGM scenario

requirements without confounding factors from feature set differences.

C. Feature Sets and Experimental Design

To quantify the relative contribution of different data modalities, we conducted three controlled ablation experiments. Each experiment systematically removed feature groups while maintaining consistent model training and evaluation procedures, enabling direct assessment of each modality's predictive value through performance-degradation analysis.

Experiment 1 (All Features) served as the baseline, utilizing the complete feature set described in Section II.B. This configuration combines glucose trajectory features (lags, EMA, slope), meal composition (macronutrients), temporal indicators (cyclical time encoding, meal period flags), and participant diagnosis status. The all-features configuration was evaluated across all four observation windows (W15, W30, W45, W60) to establish the upper bound of performance for each temporal context.

Experiment 2 (Glucose-Only) isolated physiological signals by retaining exclusively glucose-derived features: the lagged measurements (*GL_t-X*), the exponential moving average (*GL_EMA*), and the linear slope (*GL_slope*). All meal composition, temporal, and demographic features were excluded. This experiment tests whether recent glucose dynamics alone provide sufficient information for short-term postprandial prediction, or whether exogenous features substantially improve accuracy. The glucose-only configuration was similarly evaluated across all four window lengths to assess whether the importance of meal information varies with observation duration.

Experiment 3 (Meal-Only) evaluated the converse scenario, incorporating only meal macronutrients (carbohydrates, protein, fat, fiber, calories), temporal features, and diagnosis status while excluding all glucose measurements. This configuration represents a pre-meal prediction scenario where no CGM data are available, testing whether nutritional composition and timing alone can reliably forecast the postprandial glucose response. Since meal composition features are static and do not depend on the pre-meal observation window length, this experiment was conducted using the dataset structure aligned with the validation folds of the other experiments. Note that no time-series glucose data were included in this configuration; the 'window' reference purely denotes the exclusion of glucose history. Table 2 summarizes the experimental design. All three experiments employed identical training procedures: Random Forest, XGBoost, and CatBoost algorithms with class weighting to address the 3.06:1 imbalance ratio, trained via 5-fold group cross-validation based on participant IDs. This ablation framework enabled isolation of each feature modality's contribution while controlling for algorithmic and validation differences. Data leakage was prevented by excluding participant identifiers, meal timestamps, exact event times, and any future glucose measurements beyond the 60-minute prediction horizon from all feature sets.

D. Machine Learning Models

Three tree-based ensemble algorithms were selected for comparative evaluation: Random Forest, XGBoost, and CatBoost. These models were chosen over distance-based algorithms (e.g., SVMs, k-NN) and Neural Networks for several reasons specific to structured biomedical data. First, tree-based ensembles are invariant to feature scaling and robust to the non-linear interactions between physiological signals and meal timing. Second, while Deep Learning excels in perceptual tasks, Gradient-Boosted Decision Trees (GBDT) consistently outperform Deep Learning on moderate-sized tabular datasets [10][11][12]. This study specifically compares Random Forest (a bagging technique) against XGBoost and CatBoost (boosting techniques) to evaluate the trade-off between variance reduction and bias reduction. Bagging builds independent trees to reduce variance, making it robust against noise, whereas boosting builds trees sequentially to correct prior errors, focusing on minimizing bias. All models were trained with class weighting to address the 3.06:1 imbalance ratio without synthetic data generation, which could introduce artifacts in time-series data [21].

1. Random Forest

Random Forest constructs an ensemble of decision trees trained on different random samples of data with randomly selected feature subsets [26][27][28]. Each tree is built using bootstrap sampling and evaluates

Manuscript 10 November 2025; Revised 5 January 2026; Accepted 12 January 2026; Available online 17 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1401

Table 3. Model hyperparameter settings

Parameter	Random Forest	XGBoost	CatBoost
Trees/Iterations	n_estimators=100	n_estimators=100	iterations=1000
Tree Depth	max_depth=None	max_depth=6	depth=6
Learning Rate	N/A	learning_rate=0.3	learning_rate=0.03
Regularization	N/A	N/A	l2_leaf_reg=3
Sampling	min_samples_split=2, min_samples_leaf=1	subsample=1.0, colsample_bytree=1.0	subsample=1.0, colsample_bylevel=1.0
Class Imbalance	class_weight='balanced'	scale_pos_weight=3.06	auto_class_weights='Balanced'
Loss Function	gini (default)	logloss (default)	Logloss
Feature Selection	max_features='sqrt'	N/A	N/A

only a subset of available features at each split [29]. Final predictions aggregate individual tree outputs through majority voting as in Eq. (6) [29]:

$$\hat{y} = \text{mode}\{h_1(x), h_2(x), \ldots, h_T(x)\} \quad (6)$$

where $h_t(\mathbf{x})$represents the prediction of the t-th tree and $T = 100$is the total number of trees. The ensemble approach reduces overfitting risk compared to single decision trees while maintaining computational efficiency and resistance to outliers [30][31]. Class weighting ('balanced') adjusts misclassification costs to emphasize the minority hyperglycemia class.

2. XGBoost

XGBoost (Extreme Gradient Boosting) implements gradient boosting with decision trees, iteratively refining model boundaries to minimize loss functions [32]. The algorithm constructs the ensemble sequentially by adding weak learners as shown in Eq. (7) [32]:

$$\mathrm{F_m(X)} = \mathrm{F_{m-1}(X)} + \eta \cdot \mathrm{f_m(X)} \quad (7)$$

where $F_m(\mathbf{X})$ represents the ensemble prediction at iteration m, $f_m(\mathbf{X})$is the newly added weak learner, and $\eta = 0.3$ is the learning rate that controls step size and model complexity. The algorithm employs regularization techniques to reduce model complexity and prevent overfitting while maintaining superior predictive performance [33]. XGBoost handles sparse data and missing values effectively without extensive preprocessing, and supports parallel processing for computational efficiency [34]. The scale_pos_weight parameter (set to 3.06) addresses class imbalance by increasing the penalty for misclassifying minority class samples.

3. CatBoost

CatBoost (Categorical Boosting) employs ordered boosting and specialized categorical feature handling to eliminate prediction shift caused by target leakage in traditional gradient boosting [35]. The algorithm uses random permutations of the training data to compute unbiased gradient estimates. The formula is like in Eq. (8) [35].

$$\widehat{y_i} = \sum_{j=1}^{m} h_j\left(x_i; \sigma_j\right) \quad (8)$$

where σ_j represents a random permutation of the training instances at iteration j, ensuring that each model in the sequence is trained on data that has not been used to compute the target statistics. The algorithm processes categorical variables natively without manual encoding, capturing complex non-linear relationships while maintaining model interpretability [36]. CatBoost's auto_class_weights parameter, set to ('Balanced' mode) adjusts for imbalanced outcomes without requiring manual ratio calculation. The framework has demonstrated strong performance across diverse applications, including medical diagnosis and risk assessment [18]. Table 3 summarizes the hyperparameter configurations for all three algorithms. Default settings were used for most parameters to establish a baseline performance without extensive tuning, with class weighting as the primary strategy for mitigating imbalance.

E. Experimental Setup and Evaluation Metrics

The models were rigorously validated using 5-Fold Group Cross-Validation (GCV). This approach is critical in time-series and physiological data analysis, as it ensures that all data points from a single participant are retained within the same fold (training or testing), preventing data leakage and providing a more realistic estimate of the model's generalization performance on unseen individuals [37]. The folds were grouped by *participant_id*. This grouping strategy ensures that all meal events from a specific individual appear exclusively in either the training or the test set, but not in both. This strictly prevents data leakage and

Manuscript 10 November 2025; Revised 5 January 2026; Accepted 12 January 2026; Available online 17 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1401

simulates a real-world clinical scenario in which the model must generalize to a new patient without having been trained on their specific historical data.

To evaluate classification performance, the confusion matrix was used as a fundamental tool. It categorizes prediction outcomes into four types. True Positive (TP) and True Negative (TN) represent correct classifications, and False Positive (FP) and False Negative (FN) represent classification errors [26]. Based on these categories, four primary evaluation metrics were calculated using the following equations.

a. Accuracy determines the percentage of correctly classified instances relative to the total number of samples in the dataset, with the formula as in Eq. (9) [31].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (9)$$

b. Precision evaluates the reliability of positive classifications by calculating the ratio of true positives to the total number of instances predicted as positive, with the formula as in Eq. (10) [31].

$$Precision = \frac{TP}{TP+FP} \quad (10)$$

c. Recall, also known as Sensitivity, assesses the model's effectiveness in identifying all actual positive instances, with the formula as in Eq. (11) [31].

$$Recall = \frac{TP}{TP+FN} \quad (11)$$

d. The F1 Score represents the harmonic mean of precision and recall. It provides a comprehensive view of the trade-off between the two metrics, particularly when dealing with class-imbalanced datasets, with the formula as in Eq. (12) [31].

$$F1-Score = 2 \times \frac{Precission \times Recall}{Precission+Recall} \quad (12)$$

For this specific analysis, model performance was evaluated using metrics particularly relevant to imbalanced classification and clinical utility [21]. These included the F1-Macro Score, which provides a balanced measure across both normal and hyperglycemic classes; Recall for Class 1 (hyperglycemic events), a crucial clinical metric to minimize false negatives; overall Accuracy; and Precision for Class 1. The model with the highest F1-Macro score was selected as the best-performing model. This study comprehensively evaluated 27 configurations combining three feature sets, four input window lengths, and three classification algorithms. All experiments were conducted in Python using the scikit-learn, CatBoost, and XGBoost libraries, with reproducibility ensured by a fixed random seed of 42.

To ensure reproducibility, all experiments were conducted with a consistent computational environment using Python 3.11.5. Key libraries included scikit-learn (v1.4.2), xgboost (v3.1.0), and catboost (v1.2.8), and the training process was performed on a Kaggle notebook.

III. Result

A. Experiment 1: All Features

Experiment 1 evaluated the complete feature set. Table 4 presents the results of postprandial hyperglycemia classification using all features. Across all models and windows, the CatBoost model with a 30-minute window (W30) achieved the highest overall performance, with an F1-Macro of 0.7263 and an overall accuracy of 0.8155. This model also achieved the highest recall for the hyperglycemia class (Class 1) at 0.6404.

Table 4. Machine learning classification models' performance on experiment 1

Window	Model	Acc	F1-Macro	Rec
W15	Random Forest	0.805	0.690	0.427
	XGBoost	0.794	0.703	0.528
	Catboost	0.785	**0.715**	0.640
W30	Random Forest	0.807	0.684	0.404
	XGBoost	0.793	0.697	0.507
	Catboost	0.796	**0.726**	0.640
W45	Random Forest	0.807	0.683	0.392
	XGBoost	0.796	0.700	0.506
	Catboost	0.796	**0.722**	0.621
W60	Random Forest	0.797	0.661	0.357
	XGBoost	0.795	0.703	0.518
	Catboost	0.796	**0.720**	0.609

However, the advantage over a 15-minute window (W15) was modest. CatBoost showed minimal performance degradation with shorter windows, with F1-Macro dropping only 1 percentage point from W30 to W15. When F1-Macro is averaged across the three algorithms, W15 and W30 become practically indistinguishable. A 15-minute pre-meal history already captures the most informative CGM dynamics. Extending the window to 30 minutes provides only a small performance gain at the cost of doubling the required history for real-time deployment. In contrast, further extending the input history to 45 or 60 minutes did not yield consistent improvements and, in some cases, slightly degraded performance, suggesting diminishing returns beyond 30 minutes. These results indicate that W30 is the performance-optimal choice, whereas W15 can be regarded as a near-minimal input window that preserves most of the discriminative power while minimizing latency and buffering requirements.

Manuscript 10 November 2025; Revised 5 January 2026; Accepted 12 January 2026; Available online 17 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1401

B. Experiment 2: Glucose-Only Features

Removing meal and temporal data moderately reduced performance, as shown in Table 5. The best Glucose-Only model (CatBoost W30) achieved an F1-Macro of 0.6848. The Recall_1 for CatBoost W30 in Experiment 2 (0.6539) matched Experiment 1. However, Precision declined significantly (0.4666 in Exp 2 vs. 0.5394 in Exp 1 for Class 1), resulting in more false positives. Meal data appears to help distinguish true physiological spikes from transient noise. A similar but slightly stronger pattern was observed in the Glucose-Only setting: W30 systematically improved F1-Macro by around 2 percentage points over W15 across all three algorithms, while Recall_1 increased by less than 0.03. This reinforces that most of the predictive signal is already contained in the first 15 minutes of pre-meal CGM history, with 30 minutes providing only incremental refinements.

Table 5. Machine learning classification models' performance on experiment 2

Window	Model	Acc	F1-Macro	Rec
W15	Random Forest	0.750	0.602	0.299
	XGBoost	0.721	0.620	0.445
	Catboost	0.742	**0.673**	0.646
W30	Random Forest	0.769	0.624	0.328
	XGBoost	0.748	0.641	0.443
	Catboost	0.752	**0.685**	0.654
W45	Random Forest	0.778	0.639	0.345
	XGBoost	0.750	0.642	0.441
	Catboost	0.756	**0.679**	0.601
W60	Random Forest	0.798	0.667	0.379
	XGBoost	0.761	0.656	0.454
	Catboost	0.756	**0.679**	0.602

C. Experiment 3: Meal-Only Features

Models that relied exclusively on meal composition and timing data demonstrated substantially inferior performance, with the best-performing model (Random Forest) achieving an F1-Macro score of only 0.5937. Hyperglycemia recall reached only 0.42, indicating that nutritional data alone cannot provide reliable hyperglycemia risk monitoring without concurrent physiological measurements. This underscores the necessity of incorporating real-time glucose measurements for clinical safety applications. Table 6 presents the classification performance on the Meal-Only feature set. Feature ablation with W15 as baseline quantified the relative importance of each modality. Removing meal features from the complete feature set led to a 5.9% F1-Macro decline in CatBoost and a 12.8% decline in Random Forest, while removing glucose history resulted in an even more dramatic 18.8% F1-Macro reduction for CatBoost. These findings confirm that glucose history constitutes the primary signal for hyperglycemia prediction, while meal information provides supplementary value (6-13% improvement) that enhances clinical reliability. The best performing classification model for all experiments is shown in Fig. 2.

Table 6. Machine learning classification models' performance on experiment 3

Model	Acc	F1-Macro	Rec
Random Forest	0.695	**0.594**	0.423
XGBoost	0.644	0.572	0.508
Catboost	0.648	0.581	0.539

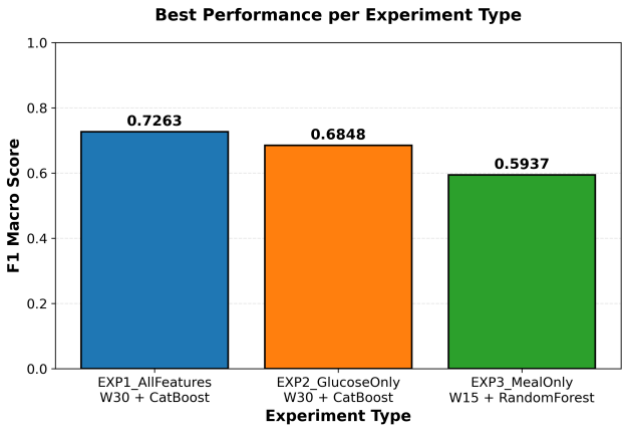

Fig. 2. The best model performance for every experiment setting

IV. Discussion

This study compared observation windows and feature sets for predicting postprandial hyperglycemia. The best configuration used a 30-minute pre-meal window (W30) with CatBoost and all features, achieving an F1-Macro of 0.7263, an accuracy of approximately 0.80, and a hyperglycemia recall of approximately 0.64. Importantly, the 15-minute window (W15) yielded only a slight reduction in F1-Macro (0.7154 for CatBoost). To rigorously validate these performance differences, we conducted paired t-tests across the 5 validation folds. Comparing the best-performing window (W30) against the minimal window (W15) for CatBoost revealed no statistically significant difference ($p=0.13 > 0.05$). This statistical equivalence is a crucial finding: it empirically demonstrates that a 15-minute pre-meal history captures most of the discriminative predictive signal. Longer windows (45 and 60 minutes) offered no systematic advantage and occasionally degraded performance. This phenomenon likely stems from "temporal noise," where the inclusion of glucose data from 45–60 minutes prior introduces physiological states that are less correlated with the immediate post-meal response, increasing feature dimensionality and

Manuscript 10 November 2025; Revised 5 January 2026; Accepted 12 January 2026; Available online 17 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1401

overfitting risk without adding informative signal. This introduces a clear trade-off between accuracy and latency. A 30-minute window is preferable when slightly higher F1-Macro and stability across models are prioritized. A 15-minute window suits real-time or resource-constrained applications, for example, during CGM warm-up, in the presence of missing data, or on embedded devices, because it offers comparable sensitivity to hyperglycemia using only half as much pre-meal history. Adopting W15 may be justified when a 1-3 percentage point F1-Macro reduction is acceptable in exchange for faster and lighter-weight prediction. From a system engineering perspective, reducing the required input history from 60 to 30 minutes translates into tangible benefits for embedded implementation. It effectively halves the memory allocation required for data buffering on microcontrollers and reduces the computational complexity of feature extraction. For battery-powered wearable devices, this reduction can lower the processor's duty cycle and reduce Bluetooth transmission payloads, thereby extending device operating life.

Feature ablation confirmed that CGM history dominates short-term postprandial prediction. Glucose-only models retained most of the discriminative performance of the full models (approx. 94%). However, the 'All Features' configuration, which integrated glucose history with meal composition, temporal context (time of day), and participant diagnosis, provided a consistent performance boost (approx. 4% in F1-Macro). This suggests that while glucose dynamics capture the immediate physiological state, non-glucose features provide critical contextual constraints. Specifically, temporal features likely help the model account for circadian variations in insulin sensitivity (e.g., morning vs. evening resistance), while diagnosis status sets a baseline risk probability for each user. In contrast, meal-only models using macronutrient composition and time-of-day features performed substantially worse, with the best model achieving an F1-Macro of approximately 0.59 and a hyperglycemia recall of approximately 0.42. This pattern aligns with other CGM-driven risk models, where CGM-derived indices alone achieve high discrimination [38][39]. Additional clinical or demographic features provide only modest incremental gain. In the postprandial setting, meal information improves precision and F1-Macro by approximately 6-13%, but it is insufficient on its own for safety-relevant prediction. However, despite its lower accuracy, the Meal-Only model holds potential for specific 'cold-start' scenarios, such as newly diagnosed patients who have not yet been prescribed a CGM, or in resource-constrained settings where CGM sensors are unaffordable. In these cases, a prediction based solely on dietary intake, while less precise, provides a baseline educational tool to estimate post-meal risk better than random chance.

These findings align with recent comparisons of multivariate and univariate glucose forecasting. Nemat et al. reported that simply appending carbohydrate, insulin, and physical activity variables to CGM inputs did not systematically improve forecasting accuracy, and, in some cases, worsened performance, indicating that multivariate models require careful data fusion to extract value from additional modalities [40]. Simple concatenation of feature vectors often fails to capture the complex interdependencies between static inputs (meal composition) and dynamic time-series (glucose trends). While "careful fusion" in deep learning contexts often entails advanced architectures like multi-stream networks to align these modalities, our results suggest that for structured biomedical data, gradient-boosted trees offer a robust alternative. Tree-based models inherently handle this fusion by learning non-linear interactions between static and dynamic features at decision-split nodes, effectively integrating the modalities without the need for complex architectural engineering. Seo et al. used Random Forest models with CGM-derived features and mealtime announcements to predict 30-minute postprandial hypoglycemia and showed that adding structured, event-aligned information can improve performance in that risk domain [41]. In the present setting, meal composition and time-of-day features, when combined with pre-meal CGM history, yielded measurable gains in F1-Macro and reduced false positives, but without concurrent CGM measurements, they did not provide a sufficiently reliable signal for hyperglycemia risk estimation.

This work complements prior postprandial-specific modeling that has focused either on hypoglycemia alone or on longer input windows and mechanistic features. Cui et al. proposed a unified LSTM model that jointly predicts postprandial hyperglycemia and hypoglycemia using a one-hour CGM history and achieved strong classification performance on the OhioT1DM dataset, highlighting the benefit of treating the postprandial regime separately from nocturnal or fasting periods because of distributional shifts in glucose trajectories [42]. The present analysis extends this approach to shorter pre-meal windows and traditional gradient-boosted tree models on a different dataset that includes explicit meal composition. Lifestyle-centric frameworks such as GlucoLens integrate CGM with detailed food logs and wearable-derived activity measures to predict postprandial AUC and hyperglycemia and to generate counterfactual explanations for behavioral interventions [43]. Together with prospective protocols that model individual postprandial responses from diet, CGM, and rich

Manuscript 10 November 2025; Revised 5 January 2026; Accepted 12 January 2026; Available online 17 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1401

clinical covariates [44], these studies suggest that meal and behavioral variables play a larger role in long-term AUC or personalized dietary guidance. For 60-minute hyperglycemia prediction, CGM history remains the dominant predictor.

The most clinically relevant metric, hyperglycemia recall, reached approximately two-thirds in the best configuration, with an acceptable overall F1-Macro despite class imbalance. This performance reflects the impact of the class weighting strategy (3.06:1), which deliberately prioritized sensitivity (Recall) over precision to minimize dangerous False Negatives (missed hyperglycemia). While this trade-off results in more False Positives (alarms for events that do not occur), in a decision-support context, this is preferable to missing a high-risk event. This sensitivity level does not yet support stand-alone therapeutic automation but may enable risk stratification, triage, and patient-facing alerts that prompt closer postprandial monitoring. The finding that a 30-minute pre-meal window is sufficient and that the required inputs are limited to CGM and basic meal information facilitates integration into existing CGM platforms or mobile decision support tools. The sharp performance degradation for meal-only models indicates that prediction systems require CGM data.

Several limitations should be noted. The analysis relies on a single dataset with a fixed 60-minute prediction horizon, derived from Dexcom G6 data. Consequently, the established 30-minute optimal window is specific to this configuration. It is plausible that different prediction horizons (e.g., 30 or 90 minutes) or sensors with different lag times (e.g., Flash Glucose Monitoring) might exhibit different optimal history lengths. Future studies should examine this potential variability to determine whether the window size must scale linearly with the prediction horizon. Additionally, the study used a binary definition of hyperglycemia at 180 mg/dL without modelling episode duration or severity and without external validation. Insulin dosing, physical activity, sleep, and other behavioral factors were not considered, despite their known influence on postprandial dynamics [40][43][44]. Hyperparameter tuning was limited, and deep learning models were not evaluated.

Future work should validate these findings on independent cohorts with different dietary patterns and sensor technologies. Additional directions include exploring behavioral and clinical covariates, comparing lightweight sequence models, and extending to multi-horizon or joint hyperglycemia-hypoglycemia prediction. Incorporating explainability techniques such as SHAP (Shapley Additive exPlanations) is a critical next step for clinical adoption. Beyond black-box prediction, explainability would allow the system to inform the patient why a spike is predicted, for instance, by distinguishing whether the risk is driven by the specific meal composition (high carb) or the time of day (morning resistance). This transparency would empower patients to make targeted behavioral adjustments rather than reacting to generic alerts.

V. Conclusion

This study systematically evaluated how observation window length and feature composition affect machine-learning predictions of postprandial hyperglycemia. Through controlled ablation experiments on 1,642 meal events from 45 participants, we established three key findings. First, short pre-meal observation windows yield performance comparable to that of extended histories. The optimal 30-minute window (72.6% F1-macro, 64.0% recall) outperformed the 15-minute window by only 1.5 percentage points, while 45- and 60-minute windows provided no consistent advantage. This demonstrates that recent glucose dynamics capture sufficient information for reliable 60-minute-ahead prediction, enabling low-latency systems with minimal buffering requirements. Second, systematic feature ablation revealed that glucose-trajectory features dominate predictive performance. Glucose-only models retained 94% of full model performance (68.5% F1-macro), whereas meal-only models achieved just 82% (59.4% F1-macro). This quantifies the relative contributions of different data modalities and indicates that glucose history is essential, whereas meal composition provides complementary but secondary value. Third, this study employed rigorous group cross-validation, separating participants between training and test sets, ensuring predictions generalize to unseen individuals rather than exploiting person-specific patterns. This validation strategy provides realistic performance estimates for deployment scenarios where systems encounter new users. Combined with our window and feature findings, these results yield actionable guidelines: 15-30-minute observation windows balance accuracy and responsiveness; glucose features are essential; and meal features enhance but are not sufficient on their own. Future work should validate these patterns across sensors and extend to multi-horizon prediction with explainable AI methods.

Acknowledgment

The authors would like to express their gratitude to the Department of Computer Science at Lambung Mangkurat University for its support and resources provided throughout this research.

Funding

This work was funded by the DRTPM Research Program, Ministry of Higher Education, Science, and

Manuscript 10 November 2025; Revised 5 January 2026; Accepted 12 January 2026; Available online 17 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1401

Technology of Indonesia (Main Contract No. 075/C3/DT.05.00/PL/2025; Derivative Contract No. 1455/UN8.2/PG/2025).

Data Availability

Original dataset is available via PhysioNet [23] or using this link https://physionet.org/content/cgmacros/1.0.0/. Preprocessed data for this study is available at https://drive.google.com/drive/folders/1te3qb-yY74mlsAmFfK7hFn59xXBAoT2x?usp=sharing.

Author Contribution

MRAM conducted data processing, data analysis, and interpretation, and wrote the manuscript draft. FI conceptualized and designed the study and revised the draft. FA, DK, and MIM contributed to data interpretation and manuscript revisions. All authors reviewed and approved the final version of the manuscript and agreed to be responsible for all aspects of the work, ensuring integrity and accuracy.

Declarations

Ethical Approval

This study uses publicly available, anonymized secondary data; therefore, ethical approval was not required.

Consent for Publication Participants.

Consent for publication was given by all participants

Competing Interests

The authors declare no competing interests.

References

[1] "IDF Diabetes Atlas 10th edition." [Online]. Available: www.diabetesatlas.org

[2] A. D. A. P. P. Committee, "2. Diagnosis and Classification of Diabetes: Standards of Care in Diabetes—2025," *Diabetes Care*, vol. 48, no. Supplement_1, pp. S27–S49, Dec. 2024, doi: 10.2337/dc25-S002.

[3] H. Ali, I. K. Niazi, D. White, M. N. Akhter, and S. Madanian, "Comparison of Machine Learning Models for Predicting Interstitial Glucose Using Smart Watch and Food Log," *Electronics (Switzerland)*, vol. 13, no. 16, Aug. 2024, doi: 10.3390/electronics13163192.

[4] A. E. S. El-Bashbishy and H. M. El-Bakry, "Pediatric diabetes prediction using deep learning," *Sci Rep*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-51438-4.

[5] T. Prioleau, A. Bartolome, R. Comi, and C. Stanger, "DiaTrend: A dataset from advanced diabetes technology to enable development of novel analytic solutions," *Sci Data*, vol. 10, no. 1, Dec. 2023, doi: 10.1038/s41597-023-02469-5.

[6] A. Gudiño-Ochoa *et al.*, "Enhanced Diabetes Detection and Blood Glucose Prediction Using TinyML-Integrated E-Nose and Breath Analysis: A Novel Approach Combining Synthetic and Real-World Data," *Bioengineering*, vol. 11, no. 11, Nov. 2024, doi: 10.3390/bioengineering11111065.

[7] L. Vu *et al.*, "Predicting Nocturnal Hypoglycemia from Continuous Glucose Monitoring Data with Extended Prediction Horizon."

[8] X. Wang, Z. Sun, H. Xue, and R. An, "Artificial Intelligence Applications to Personalized Dietary Recommendations: A Systematic Review," Jun. 01, 2025, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/healthcare13121417.

[9] B. O. Olorunfemi *et al.*, "Efficient diagnosis of diabetes mellitus using an improved ensemble method," *Sci Rep*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-87767-1.

[10] H. Seto *et al.*, "Gradient boosting decision tree becomes more reliable than logistic regression in predicting probability for diabetes with big data," *Sci Rep*, vol. 12, no. 1, Dec. 2022, doi: 10.1038/s41598-022-20149-z.

[11] B. Lu, Y. Cui, P. Belsare, C. Stanger, X. Zhou, and T. Prioleau, "Mealtime prediction using wearable insulin pump data to support diabetes management," *Sci Rep*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-71630-w.

[12] Y. Fan, "Diabetes diagnosis using a hybrid CNN LSTM MLP ensemble," *Sci Rep*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-12151-y.

[13] R. A. Fraser, R. J. Walker, J. A. Campbell, O. Ekwunife, and L. E. Egede, "Integration of artificial intelligence and wearable technology in the management of diabetes and prediabetes," *NPJ Digit Med*, vol. 8, no. 1, Dec. 2025, doi: 10.1038/s41746-025-02036-9.

[14] Q. Zhao *et al.*, "Chinese diabetes datasets for data-driven machine learning," *Sci Data*, vol. 10, no. 1, Dec. 2023, doi: 10.1038/s41597-023-01940-7.

[15] X. Xiong, Y. Xue, Y. Cai, J. He, and H. Su, "Prediction of personalised postprandial glycaemic response in type 1 diabetes mellitus," *Front Endocrinol (Lausanne)*, vol. 15, 2024, doi: 10.3389/fendo.2024.1423303.

[16] C. Mosquera-Lopez *et al.*, "Enabling fully automated insulin delivery through meal detection and size estimation using Artificial Intelligence," *NPJ Digit Med*, vol. 6, no. 1, Dec. 2023, doi: 10.1038/s41746-023-00783-1.

[17] M. S. Haleem *et al.*, "A multimodal deep learning architecture for predicting interstitial glucose for effective type 2 diabetes management," *Sci Rep*,

Manuscript 10 November 2025; Revised 5 January 2026; Accepted 12 January 2026; Available online 17 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1401

vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-07272-3.

[18] C. M. L. Hughes, Y. Zhang, A. Pourhossein, and T. Jurasova, "A comparative analysis of binary and multi-class classification machine learning algorithms to detect current frailty status using the English longitudinal study of ageing (ELSA)," *Frontiers in Aging*, vol. 6, 2025, doi: 10.3389/fragi.2025.1501168.

[19] B. Theodorou, B. Danek, V. Tummala, S. P. Kumar, B. Malin, and J. Sun, "Improving medical machine learning models with generative balancing for equity and excellence," *NPJ Digit Med*, vol. 8, no. 1, Dec. 2025, doi: 10.1038/s41746-025-01438-z.

[20] M. Salmi, D. Atif, D. Oliva, A. Abraham, and S. Ventura, "Handling imbalanced medical datasets: review of a decade of research," *Artif Intell Rev*, vol. 57, no. 10, Oct. 2024, doi: 10.1007/s10462-024-10884-2.

[21] Y. Yang, H. A. Khorshidi, and U. Aickelin, "A review on over-sampling techniques in classification of multi-class imbalanced datasets: insights for medical problems," 2024, *Frontiers Media SA*. doi: 10.3389/fdgth.2024.1430245.

[22] X. Chen, T. Wang, J. Zhou, Z. Song, X. Gao, and X. Zhang, "Evaluating and mitigating bias in AI-based medical text generation," *Nat Comput Sci*, vol. 5, pp. 388–396, Apr. 2025, doi: 10.1038/s43588-025-00789-7.

[23] A. Das *et al.*, "CGMacros: a pilot scientific dataset for personalized nutrition and diet monitoring," *Scientific Data* , vol. 12, no. 1, Dec. 2025, doi: 10.1038/s41597-025-05851-7.

[24] A. L. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet," *Circulation*, vol. 101, no. 23, Jun. 2000, doi: 10.1161/01.CIR.101.23.e215.

[25] A. D. A. P. P. Committee, "6. Glycemic Goals and Hypoglycemia: Standards of Care in Diabetes—2025," *Diabetes Care*, vol. 48, no. Supplement_1, pp. S128–S145, Dec. 2024, doi: 10.2337/dc25-S006.

[26] D. N. Jawza, M. I. Mazdadi, A. Farmadi, T. H. Saragih, D. Kartini, and V. Abdullayev, "Enhancing Diabetes Prediction Accuracy Using Random Forest and XGBoost with PSO and GA-Based Feature Selection," *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 7, no. 2, pp. 295–306, Feb. 2025, doi: 10.35882/jeeemi.v7i2.626.

[27] V. B. Liu, L. Y. Sue, O. M. Padilla, and Y. Wu, "Optimizing blood glucose predictions in type 1 diabetes patients using a stacking ensemble approach," *Endocrine and Metabolic Science*, vol. 18, Jun. 2025, doi: 10.1016/j.endmts.2025.100253.

[28] S. Dhanka and S. Maini, "Random Forest for Heart Disease Detection: A Classification Approach," in *2021 IEEE 2nd International Conference On Electrical Power and Energy Systems (ICEPES)*, IEEE, Dec. 2021, pp. 1–3. doi: 10.1109/ICEPES52894.2021.9699506.

[29] M. Maydanchi *et al.*, "A Comparative Analysis of the Machine Learning Methods for Predicting Diabetes," *Journal of Operations Intelligence*, vol. 2, no. 1, pp. 230–251, May 2024, doi: 10.31181/jopi21202421.

[30] M. Pal and S. Parija, "Prediction of Heart Diseases using Random Forest," *J Phys Conf Ser*, vol. 1817, p. 012009, Mar. 2021, doi: 10.1088/1742-6596/1817/1/012009.

[31] A. Masbakhah, U. Sa'adah, and M. Muslikh, "Heart Disease Classification Using Random Forest and Fox Algorithm as Hyperparameter Tuning," *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 7, no. 4, pp. 964–976, Oct. 2025, doi: 10.35882/jeeemi.v7i4.932.

[32] A. Nurul Puteri and A. Sumardin, "Comparison of Random Forest and XGBoost for Diabetes Classification with SHAP and LIME Interpretation," *Jurnal Teknologi Rekayasa)*, vol. 9, no. 2, 2024, doi: 10.31544/jtera.v9.i1.2024.121-130.

[33] Z. Rafie, M. S. Talab, B. E. Z. Koor, A. Garavand, C. Salehnasab, and M. Ghaderzadeh, "Leveraging XGBoost and explainable AI for accurate prediction of type 2 diabetes," *BMC Public Health*, vol. 25, no. 1, p. 3688, Dec. 2025, doi: 10.1186/s12889-025-24953-w.

[34] S. Jaiswal and P. Gupta, *Ensemble Approach: XGBoost, CATBoost, and LightGBM for Diabetes Mellitus Risk Prediction*. 2022. doi: 10.1109/ICCSEA54677.2022.9936130.

[35] H. Yang, Z. Liu, Y. Li, H. Wei, and N. Huang, "CatBoost–Bayesian Hybrid Model Adaptively Coupled with Modified Theoretical Equations for Estimating the Undrained Shear Strength of Clay," *Applied Sciences (Switzerland)*, vol. 13, no. 9, May 2023, doi: 10.3390/app13095418.

[36] M. Hamid, F. Hajjej, A. S. Alluhaidan, and N. W. bin Mannie, "Fine tuned CatBoost machine learning approach for early detection of cardiovascular disease through predictive modeling," *Sci Rep*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-13790-x.

[37] S. Bates, T. Hastie, and R. Tibshirani, "Cross-Validation: What Does It Estimate and How Well Does It Do It?," *J Am Stat Assoc*, vol. 119, no. 546, pp. 1434–1445, Apr. 2024, doi: 10.1080/01621459.2023.2197686.

[38] S. L. Cichosz, S. S. Olesen, and M. H. Jensen, "Explainable Machine-Learning Models to Predict

Manuscript 10 November 2025; Revised 5 January 2026; Accepted 12 January 2026; Available online 17 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1401

Weekly Risk of Hyperglycemia, Hypoglycemia, and Glycemic Variability in Patients With Type 1 Diabetes Based on Continuous Glucose Monitoring," *J Diabetes Sci Technol*, 2024, doi: 10.1177/19322968241286907.

[39] V. B. Berikov, O. A. Kutnenko, J. F. Semenova, and V. V. Klimontov, "Machine Learning Models for Nocturnal Hypoglycemia Prediction in Hospitalized Patients with Type 1 Diabetes," *J Pers Med*, vol. 12, no. 8, Aug. 2022, doi: 10.3390/jpm12081262.

[40] H. Nemat, H. Khadem, J. Elliott, and M. Benaissa, "Data-driven blood glucose level prediction in type 1 diabetes: a comprehensive comparative analysis," *Sci Rep*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-70277-x.

[41] W. Seo, Y. Bin Lee, S. Lee, S. M. Jin, and S. M. Park, "A machine-learning approach to predict postprandial hypoglycemia," *BMC Med Inform Decis Mak*, vol. 19, no. 1, Nov. 2019, doi: 10.1186/s12911-019-0943-4.

[42] R. Cui, C. J. Nolan, E. Daskalaki, and H. Suominen, "Jointly Predicting Postprandial Hypoglycemia and Hyperglycemia Using Continuous Glucose Monitoring Data in Type 1 Diabetes," in *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, IEEE, Jul. 2023, pp. 1–7. doi: 10.1109/EMBC40787.2023.10340094.

[43] A. Mamun *et al.*, "LLM-Powered Prediction of Hyperglycemia and Discovery of Behavioral Treatment Pathways from Wearables and Diet," *Sensors*, vol. 25, no. 17, Sep. 2025, doi: 10.3390/s25175372.

[44] N. K. Choudhry, S. Priyadarshini, J. Swamy, and M. Mehta, "Use of Machine Learning to Predict Individual Postprandial Glycemic Responses to Food Among Individuals With Type 2 Diabetes in India: Protocol for a Prospective Cohort Study," *JMIR Res Protoc*, vol. 14, 2025, doi: 10.2196/59308.

Author Biography

Muhammad Rafly Alfarizqy Maulana is a dedicated undergraduate student at Lambung Mangkurat University, pursuing studies in the Department of Computer Science since 2021. His research primarily focuses on data science, with a strong emphasis on machine learning applications in healthcare, particularly on developing predictive models for medical analysis. He is passionate about leveraging his technical and analytical skills to develop innovative, data-driven solutions that benefit society. His current research interests include the analysis of continuous glucose monitoring data and the prediction of hyperglycemia using various ensemble machine learning methods. For further details or inquiries about potential collaboration, please contact him at rafalfarizqy.ra@gmail.com.

Fatma Indriani serves as a lecturer in the Department of Computer Science at Lambung Mangkurat University. She successfully completed her undergraduate studies in the Department of Informatics at the Bandung Institute of Technology. Since 2008, she has been a dedicated faculty member in the Department of Computer Science at Lambung Mangkurat University. To enhance her expertise, she obtained a master's degree from Monash University, Australia, in 2012, and later earned a doctorate degree in Bioinformatics from Kanazawa University, Japan, in 2022. Her primary research interest is applied Data Science. Through her academic work, she is committed to advancing technology while also mentoring students and collaborating in interdisciplinary fields.

Friska Abadi completed his bachelor's degree in Computer Science from Lambung Mangkurat University in 2011. Following his graduation, he pursued advanced studies and successfully obtained his master's degree from the Department of Informatics at STIMIK Amikom, Yogyakarta, in 2016. After completing his postgraduate studies, he returned to Lambung Mangkurat University to begin his career as a lecturer in the Department of Computer Science. Currently, he holds the important position of head of the software engineering laboratory. His primary research interest revolves around the field of software engineering, where he focuses on developing innovative systems and improving technological standards to support academic and community development.

Dwi Kartini earned her Bachelor's and Master's degrees in Computer Science from the Faculty of Computer Science at Putra Indonesia "YPTK" in Padang, Indonesia. She is also a lecturer in the Department of Computer Science, where she teaches a range of courses, including linear algebra, discrete mathematics, and research methods. Her research interests focus on the applications of Artificial Intelligence and Data

Manuscript 10 November 2025; Revised 5 January 2026; Accepted 12 January 2026; Available online 17 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1401

Mining. Currently, she serves as an assistant professor in the Department of Computer Science, Faculty of Mathematics and Natural Sciences at Lambung Mangkurat University in Banjarbaru, Indonesia, and is the head of the Computer Science study program.

Muhammad Itqan Mazdadi is a lecturer in the Computer Science Study Program at Lambung Mangkurat University, South Kalimantan. He completed his undergraduate education in Computer Science at Lambung Mangkurat University in 2008-2013. After earning a bachelor's degree, he pursued a master's degree at the Indonesian Islamic University in Yogyakarta from 2013 to 2017, with a focus on Computer Science. As an educator, he not only plays a role in transferring knowledge to students, but is also active in research and development in data centers and computer networks. His contributions in this field have had a positive impact on improving the quality of education and information technology practices.

Manuscript 10 November 2025; Revised 5 January 2026; Accepted 12 January 2026; Available online 17 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1401